Privatising Privacy: Trojan Horse in Free Open Source Distributed Social Platforms

Dušan Barok
Networked Media
Piet Zwart Institute
Rotterdam, The Netherlands

May 22, 2011

Recently we have seen the arrival of distributed/federated social platforms like Diaspora, StatusNet, GNUSocial, or Thimbl. These platforms are understood as critique of social network services. They are based on premise that open source software built upon distributed database and with more advanced privacy controls will provide non-exploitative alternative to corporate social networking sites like Facebook, Twitter, or Tumblr.

I would like to question the validity of this assumption.

In order to do this, let us look at what may be the core of competition between two social web oligarchs, Facebook and Google.

There was a small and marginal conflict between these companies in November 2010. Facebook turned down Google's request for automatic import of Facebook friend list to Gmail. Google responded by blocking Facebook from one-click import of Gmail contacts.

What is striking is that they not only blocked the feature of exploiting one network to promote the other, but that they didn't want one network access the user's friend/contact lists at the other. They were protecting the friend lists.

To understand why friend list is such a valuable thing for social media companies, let us first look at how it is connected to value extraction. It had to have something to do with their main source of income – advertising, obviously.

In 2009, 97% of Google's revenue came from advertising[1], and more than 80% in case of

[1] http://investor.google.com/financial/2009/tables.html

Facebook[2] (latter number is a rough estimate, since US private companies are not obliged to publish their revenue figures). Advertisers pay for their products placements on the websites plugged into and depending on these centralised resources. Whether it is within facebook.com and google.com domains, or across the web, relevance of ads is calculated from particular user's data. In last decade, these data consisted of profile information, such as location, age, gender, languages spoken, education, workplace, relationship status, and interests, such as likes, products, or searched terms. Personalisation of online services allowed the direct marketing to bloom. Users keep on producing their online identities, these are in turn being stored in centralised databases and rented to advertisers. This probably won't change anytime soon. Advertising proved to be the major income for pre-internet mass media too, only that with personalisation it got more sophisticated. And now with socialisation it is getting even more. Owning the information about how users are related to each other, it was just a matter of time when the network providers come up with solutions to monetize the social relations too.

One of the attempts to do this was Facebook Beacon launched in 2007. It allowed Facebook to collect data about user's purchases from external partner websites and make users share information about their purchases with their friends. Technically, the service used cookies on a client side and web bugs embedded within the third-party sites. It worked without user consent, what obviously raised privacy concerns. Class action lawsuit followed and the service shut down two years later. It did not take too long to find a roundabout. After the spread of Like buttons across the web, earlier this year Facebook launched 'Sponsored Stories Ad Unit', one of 'social ad' programs using friends to advertise the products. As the website explains: "Social adverts pair an advertiser's message with social actions you have taken, such as liking a Page. You only appear in social adverts to your confirmed friends."

To have actual friends advertise products to the users sounds like a clever move. The concept of 'homophily' used in social theory confirms this. It basically says that the "people who communicate with each other are more likely to be similar to each other". Building upon this concept, Hill, Provost and Volinsky in their paper found that " 'network neighbours'– those consumers linked to a prior consumer–adopt the service [or product] at a rate 3-5 times greater than baseline groups selected by the best practices of the firm's marketing team". Users are more likely to consume products previously consumed by their friends. It is the ground assumption of the coming next wave of direct marketing: what they call 'network-based marketing', or what Wired named 'social commerce' (Rowan and Cheshire, 2011). Authors of the paper list examples of companies collecting social data:

> "eBay purchased Skype; they now have large-scale, explicit data on who talks to whom. With Gmail, Google has access to explicit networks of consumer interrelationships and is using Gmail for marketing; directed network-based marketing might be a next step. Various systems have emerged that provide explicit linkages between acquaintances (e.g., MySpace, Facebook). As more consumers create in-

[2] https://secure.wikimedia.org/wikipedia/en/wiki/Template:Facebook_revenue

> terlinked blogs, another data source arises. More generally, these results suggests that such linkage data potentially could be a sort of data considered for acquisition by many types of firms, as purchase data now are being collected routinely by many types of retail firms through loyalty cards." They close the paper with stressing the importance of social data over other personal data for marketers: "It may well be that direct communications between people is a better indicator of deep similarity than any demographic or geographic attributes." (Hill et al 2006)

This gives Facebook reason to protect its users' friends lists. They serve as the competitive advantage. How is it with Google?

Google was not successful with social applications: operation of social networking platform Orkut was fully transferred to Google Brazil (having almost 90% users living in Brazil and India), OpenSocial framework did not receive expected attention, development of Google Wave was suspended (and transferred to Apache Foundation), and Google Buzz sharing tool was largely ignored for bringing nothing new and original. Largest revenue stream still comes from its search engine. To exploit social data from users in order to feed the social ads, Google introduced Social Search in 2009. Users are encouraged to enter their usernames they have at various social network sites into their Google Profile accounts, so that the engine can give a special ranking to the relevant content generated by their friends and contacts across the social platforms, and favor it in search results. Basically, users get their search results annotated with social information in exchange for donating their social connections from across the web to Google.[3]

This is only a part of network-based marketing story. Facebook and Google are also competing in serving as the supplier of population's social data to industries, which are being developed around them. To expand the sources for advertising revenues by including the social games and applications. They needed to develop methods of delivery while preserving control over access to these social data. They came up with the concept of social graph, which is basically the sum of the friend/contact lists of all users. In May 2007, Facebook enabled developers to "build full social applications on top of the social graph, inside of Facebook."[4]

A couple of months later, internet entrepreneur and LiveJournal creator Brad Fitzpatrick published a proposal to make "the social graph a community asset, utilizing the data from all the different sites, but not depending on any company or organization as 'the' central graph owner" (Fitzpatrick 2007). Fitzpatrick was subsequently hired by Google, which proudly launched its Social Graph API in January 2008, giving graph to the public. His plan was realised, although two fundamental conditions were omitted: API has conditioned access and data are collected and hosted exclusively on Google's servers.[5] Here is what he called for in

[3] HowGoogleSocialSearchWorks,Oct2009,http://www.youtube.com/watch?v=BlpTjP6h6Ms

[4] http://www.facebook.com/press/releases.php?p=3102

[5] http://code.google.com/apis/socialgraph/, APIdataexample:https://socialgraph.googleapis.com/lookup?q=http://twitter.com/dusanson&edo=1&edi=1&fme=1&pretty=1

original proposal:

> a. Establish a non-profit and open source software (with copyrights held by the non-profit) which collects, merges, and redistributes the graphs from all other social network sites into one global aggregated graph. This is then made available to other sites (or users) via both public APIs (for small/casual users) and downloadable data dumps, with an update stream / APIs, to get iterative updates to the graph (for larger users)
>
> b. While the non-profit's servers and databases will initially be centralized, ensure that the design is such that others can run their own instances, sharing data with each other. Think 'git', not 'svn'. Then whose APIs/servers you use is up to you, as a site owner. Or run your own instance. (Fitzpatrick 2007)

Facebook responded in April 2010 by introducing Graph API, containing much larger social graph, and expanding it by objects including photos, events, and product pages.[6] While both APIs employ the open source standards for data structure and authorisation (XFN, FOAF, JSON, OAuth), the data sets are storred centrally and access to them is limited. Facebook allows third parties to access user data only upon temporary approval by user (unless user defaulted them to public), and Google provides only "publicly declared connections".

Monopolisation of social graph goes hand in hand with privacy control. Social platforms willingly operate in a linguistic greyzone between the public and private. Their business model depends on providing customers data which are unique and which they don't find elsewhere. There was another interesting event. Facebook's model of data enclosure was challenged by exponential rise of Twitter's user base, which embraces the model of controlled open space. Consequently, in December 2009 it announced opening up all profiles with photos and other social data to the whole internet, even at a price of reinventing their revenue model. Serving as a role model, Zuckerberg opened up his own profile "sharing photos of himself at parties and with his girlfriend"[7]. To his surprise, user base went on strike to demand "privacy" and pushed Facebook back to readapt their business model of secrecy and closed data, thus legitimising it.

Google followed the example. In an update to Social Search they "given you more control over how you connect accounts, and made connecting accounts more convenient. You can still connect accounts publicly on your Google profile, but now weve added a new option to connect accounts privately in your Google Account. (After all, you may not want everyone to know youre @spongebobsuperfan on Twitter)"[8]. Playing on the privacy tune, Google gives its users protection of their social profile from others.

By being offered the privacy control settings within the network, users perform their privacy

[6] http://developers.facebook.com/docs/reference/api/, APIdataexample:https://graph.facebook.com/me/friends

[7] http://www.wired.com/epicenter/2009/12/zuckerberg-facebook-privacy/

[8] http://googleblog.blogspot.com/2011/02/update-to-google-social-search.html

and voluntarily feed in the content designated solely for their peers. This creates not only "walled gardens" of closed systems, but more importantly, privacy lock-in for users who are left to demand protection of their personal data.

As a result, by creating a problematic private/public divide, the network owners are justified to take upon the role of protector, "privatise" the private data and enclose the social graph generated in this way.

Coming back to the original question, this paper asked whether the combination of open source software, distributed database and advanced privacy controls is enough to provide the non-exploitative alternative to what was just described.

The question can now be reformulated: *how shall we treat the social graph?* Shall we treat is as a public good as Fitzpatrick originally proposed? And if it is a public good, does data about who we talk to and who we friend has other use rather than for marketing and statistical analysis?

What needs to be reminded, there *are* tools for online communication and collaboration which do not know such a thing as friends list. There are mailing lists, irc, or wikis, where people are part of the group by subscribing to it, and talk and work together without being required to declare their friend or acquaintance connections to particular others.

There seem to exist two ways out of the puzzle. Either we shall not abandon the idea of federated social platforms and promote the *open* social graph, in other words social graph as a community asset. Or we shall opt out and rather focus on the critique of the very social graph and view it as the commodified social relations, a trojan horse inside "free open source distributed" social platforms.

References

[1] Berners-Lee, T (2007), "Giant Global Graph". http://dig.csail.mit.edu/breadcrumbs/node/215

[2] Brin, D (1998) "The Transparent Society", Perseus Books

[3] Fitzpatrick, B (2007) "Thoughts on the Social Graph". http://bradfitz.com/social-graph-problem/

[4] Harwood, G (2011), "Steam-Powered Census". http://www.fact.co.uk/attachments/230_steampoweredcensus_grahamharwood_21.pdf

[5] Hill, Provost, Volinsky (2006) "Network-Based Marketing: Identifying Likely Adopters via Consumer Networks". http://opimweb.wharton.upenn.edu/documents/research/NetworkedBasedMarketing-S_Hill.pdf

[6] Rowan, D, Cheshire, T (2011) "Commerce gets social: How social networks are driving what you buy", 18 January 2011, Wired. http://www.wired.co.uk/magazine/archive/2011/02/features/social-networks-drive-commerce?page=all